A Benchmark of Popular Free Energy Approaches Revealing the Inhibitors Binding to SARS-CoV2 Mpro

Son Tung Ngo,[ab*] Nguyen Minh Tam,[bc] Minh Quan Pham,[de] Trung Hai Nguyen[ab]

[a]Laboratory of Theoretical and Computational Biophysics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[b]Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[c]Computional Chemistry Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[d]Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

[e]Institute of Natural Products Chemistry, Vietnam Academy of Science and Technology, Hanoi, Vietnam

ABSTRACT: COVID-19 pandemic has killed millions of people worldwide since its outbreak in Dec 2019. The pandemic is caused by the SARS-CoV-2 virus whose main protease (Mpro) is a promising drug target since it plays a key role in viral proliferation and replication. Currently, designing an effective therapy is an urgent task, which requires accurately estimating ligand-binding free energy to the SARS-CoV-2 Mpro. However, it should be noted that the accuracy of a free energy method probably depends on the protein target. A highly accurate approach for some targets may fail to produce a reasonable correlation with experiment when a novel enzyme is considered as a drug target. Therefore, in this context, the ligand-binding affinity to SARS-CoV-2 Mpro was calculated via various approaches. The Autodock Vina (Vina) and Autodock4 (AD4) packages were manipulated to preliminary investigate the ligand-binding affinity and pose to the SARS-CoV-2 Mpro. The binding free energy was then refined using the fast pulling of ligand (FPL), linear interaction energy (LIE), molecular mechanics-Poission Boltzmann surface area (MM-PBSA), and free energy perturbation (FEP) methods. The benchmark results indicated that for docking calculations, Vina is more accurate than AD4 and for free energy methods, FEP is the most accurate followed by LIE, FPL and MM-PBSA (FEP > LIE ≈ FPL > MM-PBSA). Moreover, the binding mechanism was also revealed by atomistic simulations. The vdW interaction is the dominant factor. The residues *Thr26*, *His41*, *Ser46*, *Asn142*, *Gly143*, *Cys145*, *His164*, *Glu166*, and *Gln189* are essential elements affecting on the binding process. The benchmark probably guide for further investigations using computational approaches.

INTRODUCTION

The SARS-CoV-2, a novel coronavirus, causes severe acute respiratory syndromes and is related to millions of the death worldwide since its first outbreak in December 2019 in Wuhan, Hobei province, China.[1-4] The virus has been thought to originate from bats and can quickly transfect between humans.[5] The spreading speed is high since it is able to exist in aerosol especially.[6] Despite efforts to limit the spread of the virus, more than 80 million people were infected within a year. The outbreak of the virus effectuates COVID-19 pandemic. Therefore, the development of an effective therapy is thus much more rugent for community health. Although remdesivir was firstly approved as the anti-viral drug for treating COVID-19,[7] it is probably considered a controversial decision[8] since the drug showed disappointing trials.[9, 10] Searching an appropriate treatment for COVID-19 is accordingly great urgency.

Coronaviruses genomes, encoding more than 20 different proteins, are known as the largest RNA viruses sequence, which is approximately 26-32 kb in length.[11, 12] The SARS-CoV-2 virus forms >82% homologous RNA genome to the SARS-CoV.[1] The SARS-CoV-2 main protease (Mpro), being more than 96% identical to the SARS-CoV Mpro, is one of the most pivotal protein because of its direct involvement with the viral replication and proliferation.[12, 13] In particular, the SARS-CoV-2 Mpro sunders 11 polyproteins to polypeptides, which are served for replication and encapsulate a new virus.[12] Several probes in both computational and experimental studies were executed chasing the scheme and acquired preliminary results,[14-18] however, as mentioned above, it was a controversial decision[8] and moreover, an effective drug inhibiting SARS-CoV-2 Mpro is still unattainable.

Currently, the computer-aided drug design (CADD) is routinely employed to rapidly screen probable inhibitors for preventing the biological function of a specific enzyme.[19, 20] The time and cost for a therapy development thus decrease. In particular, the Gibbs free energy difference of the non-covalent chemical reaction between SARS-CoV-2 Mpro and their inhibitors, ΔG, can commonly be computed using molecular dynamics (MD) simulations, because it connects with the inhibition constant, K_i, an important metric disclosing the binding mechanism among biomolecules. Reliable calculation of the ΔG is one of the most pivotal issues in CADD.[21-23] Therefore, numerous approaches have

been inspired to resolve the problem.[23] In order to screen a large number of candidates, which probably up to several million compounds, the computational probe is generally operated via two steps: initial screening thousands/millions of compounds via rapid protocols such as quantitative structure activity relationship,[24] molecular docking,[25] and machine learning;[26] the ΔG was then realized using MD simulations, in which the popular free energy estimation approaches can be listed as the linear interaction energy (LIE),[27] fast of pulling ligand (FPL),[28] molecular mechanism-Poisson-Boltzmann (generalized Born) surface area (MM-PB(GB)SA),[29, 30] non-equilibrium molecular dynamics (NEMD),[31] thermodynamic integration,[32] and free energy perturbation (FEP) approaches,[33] etc. However, it should be noted that the precision and accuracy of the ligand-binding affinity approaches somehow depend on the enzyme targets.[34-40] Therefore, in this work, we benchmarked performance of docking approaches involving Vina[25] and AD4[41] applying on the SARS-CoV-2 target. Consequently, the MD simulations were then executed to investigate the dynamics of the SARS-CoV-2 + inhibitor complexes. The relaxed complexes were then used as initial conformations for probing ligand-binding affinity using four free energy schemes including FPL,[28] LIE,[27] MM-PBSA,[32, 33] and FEP.[33] The obtained observations probably guide for further investigations using computational approaches.

MATERIALS & METHODS

Structure of Receptor and Ligand

The X-Ray diffraction structure of SARS-CoV-2 Mpro was obtained from the Protein Data Bank (PDB) with the identify of 7JYC.[42] The structure of 34 ligands was taken from PubChem database[43] referring to the previous work[44-51] and 2D structure of them was reported in Table S1 of the Supplementary (SI) file.

Molecular Docking Simulations

Vina[25] and AD4[41] were manipulated to dock available inhibitors to binding cleft of SARS-CoV-2 Mpro. The binding cleft was selected as the binding region of the compound named narlaprevir.[42] In particular, the docking-grid size of the docking was chosen as 24 × 24 × 24 Å according to the previous work.[16, 17] The modeling of docking simulations was described in **Figure 1**A.
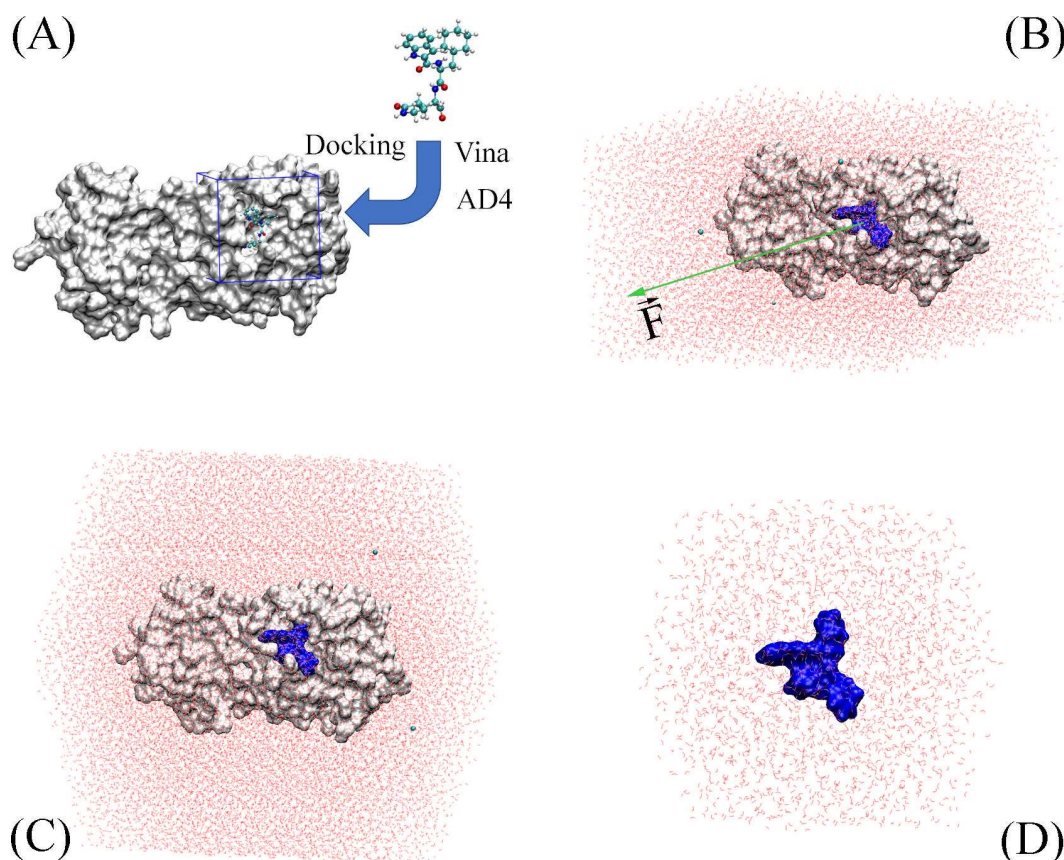
Figure 1. Computational scheme via molecular docking, SMD, and MD simulations. (A) the modeling of molecular docking simulations. Inhibitors were docked to the binding cleft of SARS-CoV-2 Mpro, which was limited in the docking box with a volume of ca. 13.82 nm³. (B) the starting structure of the SARS-CoV-2 Mpro + inhibitor for estimating the ligand-binding affinity via FPL scheme. (C) the initial shape of the SARS-CoV-2 Mpro + inhibitor for MD simulations. (D) the starting conformation of the solvated inhibitor. The cyan balls describe neutralized Na⁺ ions.

Autodock Vina (Vina).[25] The performance of Vina depends on the parameter exhaustiveness, which was chosen to be 8.[34, 52] The largest energy difference between docking shapes was set to 7 kcal mol⁻¹. The largest ligand-binding affinity was selected as the best docking structure.

Autodock4 (AD4).[41] AD4 was executed with a grid size of 72 × 72 × 72 Å with the spacing of 0.333 Å. The Autogrid4 was selected to execute the docking grid. The inhibitor was docked to the SARS-CoV-2 Mpro with the genetic algorithm (GA) run of 10, the population size of 150, and the number of

generations was 27.000. The GA number of evaluations was 250.000. The lowest binding free energy cluster was selected as the best docking conformation.

Molecular Dynamics Simulations

The MD simulations were executed to improve the docking results of available inhibitors to the SARS-CoV-2 Mpro by using GROMACS version 5.1.5.[53] In particular, the SARS-CoV-2 Mpro and neutralized ions were described via the Amber99SB-iLDN force field.[54] Consequently, the inhibitor was illustrated using general Amber force field[55] with the help of AmberTools18 and ACPYPE packages.[56, 57] It should be noted that before the ligand was parameterized, the quantum chemical calculation using B3LYP functional at 6-31G(d,p) level of theory was executed to obtain chemical information of the inhibitor. During which, the restrained electrostatic potential (RESP) approach was used to assign the atomic charges over quantum simulations using implicit solvent environment, $\varepsilon = 78.4$.[55] Moreover, the SARS-CoV-2 Mpro + inhibitor complex was placed into a rectangular or dodecahedron periodic boundary (PBC) condition box with a volume of 506 or 820 nm³ corresponding to the steered-MD (**Figure 1**B) or unbiased MD (**Figure 1**C) simulations, respectively. The soluble system hence encompasses of ca. 50 000 or 80 000 atoms, respectively, involving SARS-CoV-2 Mpro, ligand, water molecules, and neutralized ions Na⁺. Moreover, in order to carry out the perturbation simulations, the ligand was individually simulated in a dodecahedron PBC box with a volume of ca. 85 nm³ (**Figure 1**D). The soluble ligand system consists of ca. 8 000 atoms, involving ligand, water molecules, and counterbalanced ions.

The parameters for operating MD simulations were described in the previous works.[16, 17] In particular, the integral was attempted every 2 femtoseconds. A non-bonded pair was enumerated within a radius of 0.9 nm. The van der Waals (vdW) interaction was assessed using the cut-off scheme, while the electrostatic (cou) interaction was determined via the fast Particle-Mesh Ewald electrostatics scheme.[58] The solvated system was initially minimalized via the steepest descent approach. The canonical (NVT) and isobaric-isothermal (NPT) simulations, length of 0.1 and 2.0 ns, respectively, were then followed to equilibrate the system. The final conformations of NPT simulations of the solvated complex was operated as initial structure of SMD or MD simulations, which are length of 0.5 or 20.0 ns, correspondingly. Moreover, the solvated inhibitor system was run for 5.0 ns. Each system was imitated 2 times for getting better samples.

Free Energy Calculations

Fast pulling of ligand (FPL) scheme. An externally harmonic force was applied to pull the inhibitor dissociation from the SARS-CoV-2 Mpro binding cleft. The forced parameters were picked as the cantilever spring constant, $v = 600$ kJ mol⁻¹ nm⁻², and pulling velocity, $k = 0.005$ nm ps⁻¹.[28] During SMD simulations, the pulling work, W, was recorded to be able to use as a critical factor to estimate the ligand-binding affinity[28] since it correlates with the binding free energy, ΔG, via isobaric-isothermal Jarzynski equality.[59] The W is calculated as follows

$$W = v \int_0^t F(t)dt \qquad (1)$$

Linear interaction energy (LIE) calculation. The ligand-binding free energy, ΔG_{LIE}, was computed as the mean of vdW and cou interaction differences of inhibitor with its neighboring atoms over incorporation, i.e. the individual ligand in solvent (unbind state - denoted as subscript u) and the inhibitor in binding mode with the SARS-CoV-2 Mpro (bound state – specify as subscript b).

$$\Delta G_{LIE} = \alpha\left(\langle V_{l-s}^{vdW}\rangle_b - \langle V_{l-s}^{vdW}\rangle_u\right) + \beta\left(\langle V_{l-s}^{cou}\rangle_b - \langle V_{l-s}^{cou}\rangle_u\right) + \gamma \quad (2)$$

The coefficient γ, a constant, is associated with the alteration of the hydrophobic nature of the binding cleft conceding to various species of inhibitors, whereas coefficients α and β are rating parameters for nonpolar and polar interactions.[60]

Molecular mechanics – Poisson-Boltzmann surface area (MM-PBSA) analysis. The ligand-binding affinity, $\Delta G_{MM-PBSA}$, can be assessed in MD simulation via MM-PBSA approach[29, 30] as follow

$$\Delta G_{MM-PBSA} = \Delta E_{cou} + \Delta E_{vdW} + \Delta G_{sur} + \Delta G_{PB} - T\Delta S \qquad (3)$$

where ΔE_{cou}, ΔE_{vdW}, ΔG_{sur}, and ΔG_{PB} corresponds to the energetic changes in cou, vdW, nonpolar, and polar interactions, respectively; $T\Delta S$ is the entropic contribution to the $\Delta G_{MM-PBSA}$. In particular, ΔE_{cou} and ΔE_{vdW} terms were computed using GROMACS tools "gmx energy". The nonpolar metrics, ΔG_{sur}, was determined via Shrake-Rupley formula,[61] which is $\Delta G_{sur} = \gamma SASA + \beta$, where $\gamma = 0.0072$ kcal mol⁻¹ Å⁻² and $\beta = 0$.[62] The polar component, ΔG_{PB}, was assessed via numerically resolving the Poisson Boltzmann equation using an implicit solvent model.[63, 64] Finally, the entropic term can be probed via normal mode approximation.[65]

Double-annihilation binding free energy investigation. The inhibitor was changed from *bound* state to *unbound* state by using λ-alteration simulations,[66] which concur at $\lambda = 0$ and $\lambda = 1$, correspondingly. Several values of the coupling parameter λ were engaged to complete this task. The free energy change, $\Delta G_{\lambda=0\rightarrow1} = -k_B T ln \langle e^{-\frac{\Delta H}{k_B T}}\rangle_{\lambda=0}$, corresponds to the work of the ligand-annihilation process. The value can be assessed via the Bennett acceptance ratio scheme.[67] The binding free energy between SARS-CoV-2 Mpro and inhibitor, ΔG_{FEP}, is thus obtained due to the difference of the free energy changes over two annihilation-ligand processes involving demolishing inhibitor in the solvated complex, $\Delta G_{\lambda=0\rightarrow1}^{Comp}$, and inhibitor, $\Delta G_{\lambda=0\rightarrow1}^{lig}$.[22]

$$\Delta G_{FEP} = \Delta G_{\lambda=0\rightarrow1}^{Comp} - \Delta G_{\lambda=0\rightarrow1}^{lig} \qquad (4)$$

Analysis Tools

The chemicalize webserver, a tool of ChemAxon, was used to assess the protonation states of inhibitors. The Adaptive Poisson-Boltzmann Solver (APBS) webserver was executed to determine the surface charge of the protease.[64, 68] The correlation error was calculated using 1 000 rounds of the bootstrapping method.[69] The intermolecular nonbonded contact (NBC) between the ligand atoms to the residual SARS-CoV-2 Mpro was confirmed when the pair between non-hydrogen atoms of them is smaller than 4.5 Å. The intermolecular hydrogen bond (HB) between the Mpro residues and the inhibitors was endorsed when the angle ∠ acceptor (A)-hydrogen (H)-donor (D) is larger than $3\pi/4$ and the pair A-D is smaller than 3.5 Å.

RESULTS AND DISCUSSION

Molecular Docking Calculations

The obtained results were shown in Table 1 and Table S1 of the SI file. Initially, we assessed the docking results against the relevant experimental data including binding affinity and native binding poses.[44-50] The assessment includes two parts: correlation between docking and experimental ligand-binding affinity and successful-docking rate.[34] The estimated correlation coefficients for Vina and AD4 are $R_{Vina} = 0.60 \pm 0.13$ and $R_{AD4} = 0.47 \pm 0.21$, respectively. This indicates that docking energies of Vina are more strongly correlated with experiments than those of AD4. Moreover, the root-mean-square error ($RMSE$) of Vina is lower than that of AD4, namely $RMSE_{Vina} = 1.78 \pm 0.17$ and

$RMSE_{Vina} = 1.97 \pm 0.17$ kcal mol[-1], respectively. Although AD4 required much more computing resources than Vina does, its docking performances lagged behind Vina. It is probably caused by the difference in scoring functions as indicated by prior observations.[34] Furthermore, in the prior work,[70] AD4 gave poor correlation, $R = 0.36$, with the $\Delta G_{Jarzynski}$, which obtained via NEMD simulations,[31] a much more accurate free energy approach. Therefore, it may argue that Vina is the appropriate protocol for preliminary assessment of the ligand-binding affinity to the SARS-CoV-2 Mpro.

Table 1. The calculated results in comparison with the experimental values of some compounds to SARS-CoV-2 Mpro.

N°	Name	ΔG_{Vina}	ΔG_{AD4}	W	ΔG_{LIE}	$\Delta G_{MM-PBSA}$	ΔG_{FEP}	ΔG_{EXP} [a]
1	7J	-7.4	-6.0	95.7 ± 6.1	-15.04 ± 0.3	-19.3 ± 1.03	-17.95 ± 2.74	-8.69[44]
2	11a	-7.3	-8.1	109.7 ± 3.1	-14.78 ± 0.81	-29.67 ± 0.30	-18.95 ± 0.52	-9.96[45]
3	11b	-7.4	-8.0	91.3 ± 7.9	-13.99 ± 0.36	-14.41 ± 1.85	-16.53 ± 0.59	-10.13[45]
4	11r	-6.8	-6.9	96.6 ± 8.2	-15.98 ± 1.92	-15.14 ± 1.61	-20.89 ± 0.51	-9.23[46]
5	13a	-7.6	-7.6	64.7 ± 10.6	-10.07 ± 0.59	-0.71 ± 0.87	-10.94 ± 2.51	-7.70[46]
6	13b	-7.7	-7.4	81.3 ± 6.1	-16.16 ± 2.00	-19.93 ± 3.96	-16.47 ± 0.32	-8.45[46]
7	Baicalein	-6.8	-5.7	36.5 ± 8.0	-10.36 ± 2.57	-8.88 ± 2.92	-8.40 ± 2.23	-8.25[47]
8	Boceprevir	-7.1	-8.8	54.5 ± 1.8	-11.75 ± 0.85	-9.74 ± 0.48	-7.65 ± 1.31	-7.37[48]
9	Calpain Inhibitor I	-5.3	-5.4	50.2 ± 4.9	-10.77 ± 0.87	-7.51 ± 0.17	-6.41 ± 0.37	-6.94[48]
10	Calpain Inhibitor II	-5.6	-5.3	74.1 ± 22.9	-12.21 ± 0.22	-14.92 ± 6.95	-9.09 ± 2.39	-8.23[48]
11	Calpain Inhibitor XII	-6.3	-5.1	51.8 ± 5.7	-11.98 ± 0.30	-24.73 ± 1.21	-9.27 ± 0.88	-8.69[48]
12	Calpeptin	-4.9	-6.1	33.0 ± 5.4	-9.75 ± 1.07	-4.37 ± 1.89	-3.43 ± 0.97	-6.81[48]
13	Carmofur	-5.6	-6.0	39.1 ± 5.9	-9.30 ± 1.78	-3.97 ± 0.98	-7.12 ± 3.20	-7.86[49]
14	GC-373	-7.2	-6.6	53.1 ± 8.3	-12.16 ± 0.41	-12.04 ± 1.13	-10.32 ± 1.55	-8.76[50]
15	MG-115	-6.1	-5.4	57.8 ± 2.2	-11.34 ± 0.57	-8.14 ± 1.21	-9.19 ± 0.73	-7.53[48]
16	MG-132	-6.2	-5.2	71.4 ± 9.1	-11.50 ± 0.39	-12.13 ± 3.06	-8.48 ± 0.41	-7.41[48]
17	Narlaprevir	-7.5	-5.9	69.9 ± 2.1	-12.69 ± 0.05	-22.75 ± 0.25	-6.57 ± 0.50	-7.18[48]
18	PX-12	-3.9	-4.8	32.1 ± 1.0	-8.67 ± 0.05	-32.44 ± 0.73	-2.56 ± 0.30	-6.39[49]
19	Shikonin	-6.1	-6.0	27.3 ± 6.9	-9.37 ± 0.27	-0.74 ± 4.30	-3.01 ± 0.95	-6.58[49]
20	Tideglusib	-6.6	-7.1	36.5 ± 3.1	-9.92 ± 0.27	-10.56 ± 2.93	-4.26 ± 0.12	-7.95[49]

[a]The experimental binding free energies were gained based on IC50 value, approximating that the one equals to the inhibition constant K_i. The unit is kcal mol[-1].

The inhibitor-binding pose was also obtained over this process. The docking pose forms a small root-mean-square deviation (RMSD) with respect to experimental pose. It was counted as a successfully-docked conformation if the RMSD is smaller than 2 Å.[34] In particular, nine compounds including *7j*, **11a**, **11b**, **13b**, *baicalein*, *boceprevir*, *calpeptin*, GC-373 and *narlaprevir* were reported to have the experimental binding poses with the PDB IDs 6XMK,[44] 6LZE,[45] 6M0K,[45] 6Y2F,[46] 6M2N,[47] 7K40,[42] 7AKU,[71] 6WTK,[50] and 7JYC,[42] correspondingly. Over these systems, the successful-docking rate of Vina is ca. 67% with the mean RMSD of 1.97 ± 0.32 Å. It is significantly better than those by AD4 with the RMSD between docked and experimental structures of 3.22 ± 0.33 Å as represented in Figure S1 of the SI file. Therefore, it may be concluded that Vina not only formed the proper affinity results but also adopted the suitable binding pose to the SARS-CoV-2 Mpro.
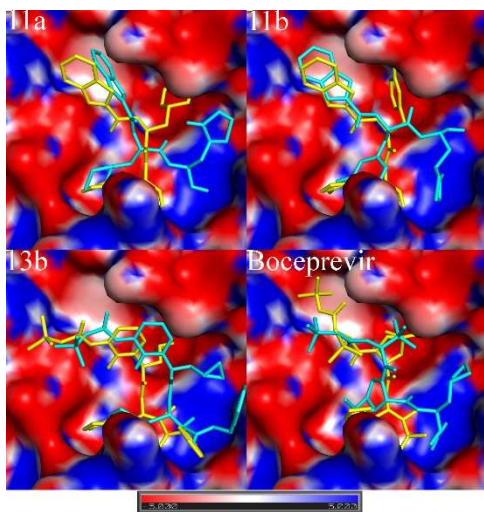
Figure 2. Comparison of docking (cyan) and experimental binding (yellow) poses of **11a**, **11b**, **13b**, and *boceprevir* to SARS-CoV-2 Mpro. The surface charge, diffuses from -5 to 5, of the protease was computed via the APBS webserver. Docking results were obtained by using Vina package.

Molecular Dynamics Simulations

Because the molecular docking simulations often use many constraints/approximations to accelerate the calculation speed, the results often need to be refined using more accurate protocols.[17, 18] In this context, because Vina formed the most suitable binding affinity and pose as discussed above, we have chosen the docking structures provided by this approach as initial structures for simulating via SMD/MD techniques. The ligand-binding free energy calculation methods were thus carried out.[23, 35] The performance of free energy calculations based on SMD/MD trajectories were thus assessed.

Steered-MD simulation. The FPL is an efficient technique to quickly classify the ligand-binding affinity.[28] This approach successfully estimated the affinities of several inhibitors binding to the SARS-CoV-2 Mpro, which suggested a shortlist of potent compounds to further evaluate via perturbation simulations.[16] A benchmark with 11 compounds were then used later on indicating that the correlation coefficient of the pulling work, W, and experimental binding free energy, ΔG_{EXP}, is appropriate with a value of $R_{FPL} = -0.76 \pm 0.10$.[17] However, due to small size of the testing set, the obtained results are probably unstable due to the large value of computed error. Consequently, the value did not show superiority over the Vina docking with $R_{Vina} = 0.72 \pm 0.14$, which is within the computed error.[17] In this context, we benchmarked again this approach for evaluating the ligand-binding affinity versus the SARS-CoV-2 Mpro with a larger testing set. The FPL scheme was thus used for refining the obtained-docking results, which were provided by Vina. In FPL simulations, an externally pulling force was applied to extract inhibitors from *bound* to *unbound* states. The recorded rupture force, F_{max}, and pulling work, W, during the simulations were given in Table S2 of the SI file. The F value in time dependence was also shown in Figure S2 of the SI file. The average of W values falls in the range 18.3 ± 1.4 to 111.3 ± 6.0 kcal mol⁻¹, providing a median of 56.0 ± 5.0 kcal mol⁻¹, while the mean rupture force F_{max} is within the range from 279.5 ± 12.7 to 1040.6 ± 68.9 pN, giving a median of 581.9 ± 41.2 pN. The

ligand-binding affinity is possibly ranked via the W value, which formed an appropriate correlation, $R = -0.51 \pm 0.15$ (cf. Table S2), with the respective experiments.[44-50] The obtained coefficient indicated that FPL is significantly worse than Vina docking, $R_{Vina} = 0.60 \pm 0.13$, in predicting the ligand-binding affinity of the SARS-CoV-2 Mpro. The poorly correlated outcomes of FPL probably appear due to the SMD simulations were performed using the conformations provided by the short NPT simulations, which may not be sufficient to reach equilibrium states. Therefore, the unbiased MD simulations with a length of 20 ns were performed after NPT simulations and were reported below. We used the last conformations of MD simulations as starting structures of FPL calculations. The obtained outcomes were reported in **Table 1** and Table S3 of the SI file, in which, the F value during SMD simulations was reported in Figure S3 of the SI file. The F_{max} and W values were thus altered and diffusing from 342.0 to 961.4 pN and 27.3 to 109.7 kcal mol⁻¹, forming median values of 644.0 ± 39.2 pN and 61.3 ± 5.3 kcal mol⁻¹, respectively. The obtained correlation between W and ΔG_{EXP} was thus increased from $R = -0.51 \pm 0.15$ to $R_{FPL} = -0.74 \pm 0.11$ (cf. **Figure 3**). The FPL technique is thus able to improve upon the docking results, however, the equilibrated simulations are required to carefully perform.
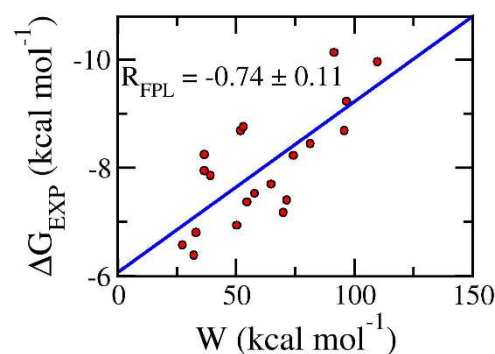
Figure 3. Association of pulling work W and ΔG_{EXP}. The W values were calculated via Eq. (1). The ΔG_{EXP} values were computed when the half-maximal inhibitory concentration, IC50, was guessed to be equal to the inhibition constant, K_i.

Unbiased MD simulation. As mentioned above, the FPL results based on the rather short relaxation time of only 2 ns were probably limited since it may not be sufficient to reach equilibrium states. Moreover, the SARS-CoV-2 Mpro C_α atoms were restrained, probably resulting in hinders the structural change of the complexed system to equilibrium states. The complexed conformation including the SARS-CoV-2 Mpro and the ligand in the best docking pose provided by Vina was thus solvated and equilibrated via unbiased MD simulations with length of 20 ns. The accuracy of the following FPL calculations was thus increased significantly (**Figure 3**). During the unbiased MD simulation, the complexes almost reached the equilibrium states after 5 ns (cf. Figure S4 of the SI file). Therefore, the snapshots over interval 10-20 ns with stride of 10 ps were collected for binding free energy calculation via the LIE and MM-PBSA approaches. Furthermore, structures extracted from MD trajectories of 2.5-5 ns of the solvated inhibitor system were also involved to free energy calculation via the LIE approach.

In order to probe the binding mechanism of inhibitors to the Mpro, the intermolecular NBC and HB between inhibitors and individual residues of the SARS-CoV2 Mro were investigated using equilibrium snapshots of all complexes. The obtained outcomes were presented in Figure S5 of the SI file, which mentioned 30 residues establishing NBC to inhibitors over more than 15 % of the appraised shapes (40 000 snapshots totally). However, there are only 19/30 residues that created HB to inhibitors. Shortening the list, we have only counted residues, which simultaneously adopted NBC and HB to inhibitors with a probability being higher than 42 and 4 %, respectively. It should be noted that 47 ± 5 and 6 ± 2 % amounts correspond to the averaged values over 30 residues. 9 residues were obtained and described in Figure 4. We may argue that the residues *Thr26, His41, Ser46, Asn142, Gly143, Cys145, His164, Glu166,* and *Gln189* are critical elements governing the binding process of ligands to the SARS-CoV-2 Mpro. Furthermore, possible mutations at these residues could change much the ligand-binding free energy to the SARS-CoV-2 Mpro.
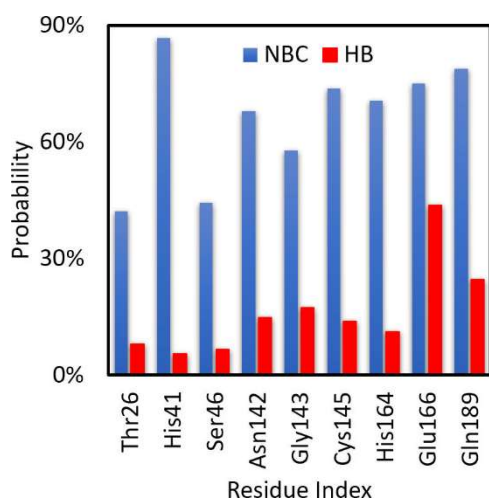

Figure 4. Critical residues forming NBC and HB to the inhibitors. The results were obtained over the equilibrium snapshots of the MD simulations of all complexes.

In addition, the clustering method was then applied to characterize the structural change of 9 critical residues during the equilibrium conformations of all complexes. The calculation was performed with the non-hydrogen atomic RMSD cutoff of 1.2 Å over 40 000 structures of 9 residues in stabilizing *bound* states with 20 inhibitors. One cluster was found, which are shown as colorful residues in Figure 5. The representative structure of 9 critical residues was compared with the starting conformation, which is in gray color. The differences between the MD refined and starting structures are noted as red arrows in Figure 5. The significant structural changes are the flexing of the residue *Asn142* and the rotation of the hydroxyl and thiol sidechains of the *Ser46* and *Cys145*, respectively. The sidechain residues probably rotate to form HB to inhibitors. Moreover, overall, the difference between the representative structure and initial conformation only is ca. 1.0 Å implying the stability of the SARS-CoV-2 Mpro active site during the MD simulations.
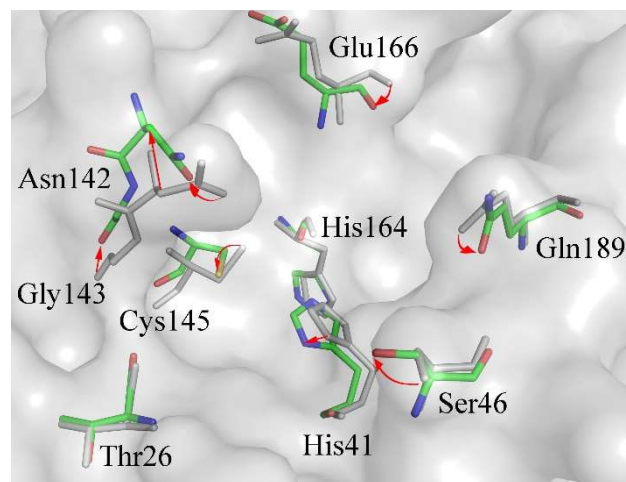

Figure 5. The representative structures of 9 critical residues via the non-hydrogen RMSD clustering calculation with a cutoff of 1.2 Å. The colorful residues represent the MD refined structure in comparison with the initial structure, which denoted using gray color. Red arrows imply the change of these residues during MD simulations.

Binding free energy calculation by using LIE scheme. The difference between averaged vdW and cou interaction energies between each inhibitor to the SARS-CoV-2 Mpro, *bound* state, and solution, *unbound* state, as long as the ΔG_{EXP} was given in Table 1 and Table S4 of the SI file. The binding free energy, ΔG_{LIE}, is computed using Eq. (2). Traditionally, the parameter α and β were chosen as 0.18 and 0.50, respectively.[27, 72] However, similar to the Aβ oligomeric system,[73] no correlation, $R = -0.13 \pm 0.20$, was observed between calculated and experimental values. This is probably due to the shallow binding cleft of the SARS-CoV-2 Mpro, which is similar to ligand-surface binding in the case of the Aβ oligomer.[73] Therefore, the parameters including $\alpha = 0.288$, $\beta = -0.049$, and $\gamma = -5.880$ of Aβ system[73] were proposed to use for calculating the ligand-binding free energy of the SARS-CoV-2 Mpro + inhibitor. Interestingly, the set of parameters gave a correlation coefficient $R_{LIE} = 0.73 \pm 0.09$ and $RMSE = 4.12 \pm 0.40$ kcal mol⁻¹ (Figure 6). Absolutely, the LIE approach formed similar accuracy outcomes, $R_{LIE} = 0.73 \pm 0.09$, compared to FPL simulations, $R_{FPL} = -0.74 \pm 0.11$. Moreover, the negative parameter β may imply the loss of cou interactions of inhibitors upon association (cf. Table S4) or it may argue that the vdW interactions control the binding process of inhibitors to the protease. It is in good agreement with the previous outcomes[16, 18] and obtained results via MM-PBSA and perturbation calculations below. Furthermore, the negative value γ implies that the hydrophobic interactions between inhibitors and the SARS-CoV-2 Mpro are strong as mentioned as conclusion about the superior of vdW term above. In addition, although the LIE adopted a good Pearson correlation, the ΔG_{LIE} overestimates the ΔG_{EXP} with an amount of ca. 3.89 kcal mol⁻¹ (see Table 1). It is probably caused by the lower hydrophobic interaction between SARS-CoV-2 Mpro + inhibitors versus Aβ complexed system or the incorrect imitation of the interaction between inhibitors and surrounding atoms.[74, 75] Overall, it may argue that the binding mechanism of the SARS-CoV-2 Mpro + inhibitor is similar to the Aβ oligomer + ligand, but the hydrophobic contacts of the Mpro complex are weaker than the Aβ ones.
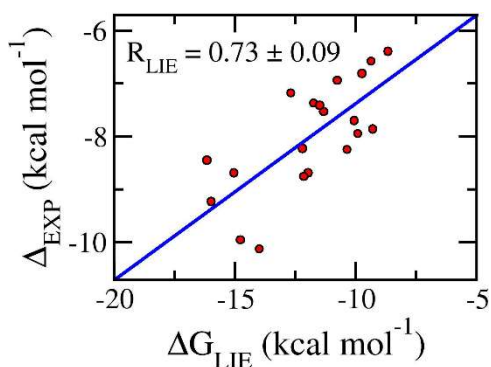
Figure 6. Comparison of ΔG_{LIE} and ΔG_{EXP}. The calculated binding free energy was computed using Eq. (2) with the parameters $\alpha = 0.288$, $\beta = -0.049$, and $\gamma = -5.880$ referring the Aβ oligomer + inhibitor.[73] The ΔG_{EXP} values were computed when the half-maximal inhibitory concentration, IC50, was guessed to be equal to the inhibition constant, K_i.

Establishing the ligand-binding free energy via MM-PBSA protocol. The equilibrium conformations of the complex SARS-CoV-2 Mpro + inhibitor during MD simulations were implemented for estimating the binding free energy using continuum models[29, 30] as follows Eq. (3). It should be highlighted that our group has been successfully calculated the ligand-binding free energy for various biomolecules using the MM-PBSA method.[73, 76-78] The obtained outcomes were described in Table 1 and Table S5 of the SI file. In particular, the $\Delta G_{MM-PBSA}$ overestimates the ΔG_{EXP} with a value of ca. 5.60 kcal mol⁻¹, which is slightly larger than that given by LIE protocol. Moreover, the MM-PBSA method provides a poor accuracy in comparison with the corresponding experiments, $R_{MM-PBSA} = 0.32 \pm 0.29$ and $RMSE = 10.15 \pm 1.92$ kcal mol⁻¹ (**Figure 7**). It is in good agreement with the previous study,[79] that MM-PBSA formed a correlation with the experiment with a value of $R_{MM-PB} = 0.25$ over investigating 15 complexes. Interestingly, as mentioned above that the binding mechanism of inhibitors to SARS-CoV-2 Mpro is quite similar to inhibitors to Aβ oligomer, the Pearson correlations of two systems are similar, $R_{MM-PBS}^{SARS-CoV-2} = 0.32$ versus $R_{MM-PB}^{A\beta} = 0.27$.[73] The poor accuracy of the MM-PBSA approach applying on the SARS-CoV-2 Mpro possibly is similar to the Aβ system that is probably caused by the selection of inappropriate dielectric constant, ε, and roughly entropic approximation.[35, 73, 80] Furthermore, the ε issue was also consolidated via the inhibitor interaction diagram analysis (cf. Tables S2 of the SI file) where the solvation exposure of inhibitors is absolutely complicate. Therefore, further investigation to characterize factors affecting the accuracy of the MM-PBSA applying on the SARS-CoV-2 Mpro should be performed before the approach would be widely used for screening potential inhibitors for the Mpro target.
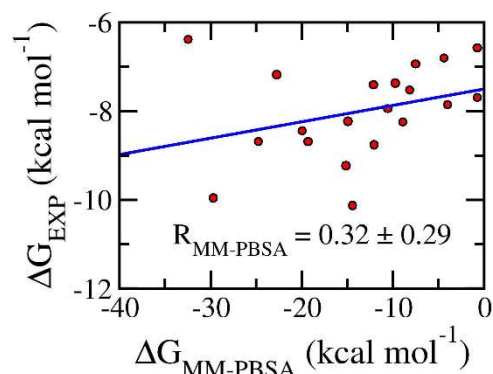
Figure 7. Comparison of $\Delta G_{MM-PBSA}$ and ΔG_{EXP}. The calculated binding free energy was computed using Eq. (3). The ΔG_{EXP} values were computed when the half-maximal inhibitory concentration, IC50, was guessed to be equal to the inhibition constant, K_i.

Determination of ligand-binding free energy using FEP method. In recent reports,[18, 79] the FEP simulations successfully determined the ligand-binding free energy as known as the most accurate free energy methods.[23, 81] However, although the perturbation results correlate with the respective experiments,[18, 79] the Pearson coefficient diffused in a large range from 0.54 to 0.94. In particular, FEP simulations determined the ligand-binding free energy of 11 inhibitors to SARS-CoV-2 Mpro with high accuracy, $R_{FEP} = 0.94 \pm 0.04$.[18] In a different study, perturbation simulations also formed a Pearson correlation $R_{FEP} = 0.54$, when 15 complexes were considered.[79] Therefore, in this work, we benchmarked the FEP performance on a larger set from multi-sources would probably provide a clarification for the accomplishment of this approach.

The final structures of MD simulations mimicking the solvated complex and ligand were utilized as input of λ-alteration simulations. The obtained results were detailly reported in Table 1 and Table S6 of the SI file. The perturbation simulations provide the highest accuracy results with a Pearson correlation of $R_{FEP} = 0.85 \pm 0.06$ (cf. **Figure 8**). The inaccurate outcomes probably caused by the unable covalent-interaction mimicking of conventional MD simulations. Moreover, in average over complexes, the ΔG_{FEP} is -9.87 ± 1.20 kcal mol⁻¹, which overestimates ca. 1.87 kcal mol⁻¹ compared to the mean of experimental values. The difference is significantly smaller than those by LIE, ca. 3.89 kcal mol⁻¹, and MM-PBSA, 5.60 kcal mol⁻¹, methods. The difference between the mean of experimental and computational values probably comes from the incorrected simulations of the interaction between inhibitors and neighboring atoms.[74, 75] The rough assumption of the IC50 equals the inhibition constant K_i, when calculated the experimental binding free energy, also adopts a shifting possibility. Furthermore, the obtained results by λ-alteration simulations also revealed the binding mechanism of the SARS-CoV-2 Mpro inhibitor. The vdW interaction is dominant in the binding process of ligands to the Mpro, which is in good agreement with the previous observations.,[16, 18] due to the average of ΔG_{cou} and ΔG_{vdW} are -2.82 ± 0.83 and -7.05 ± 0.49 kcal mol⁻¹, correspondingly.
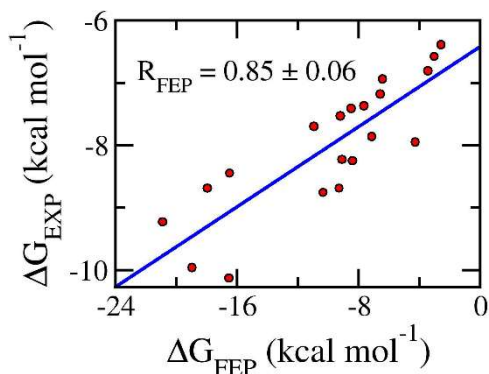
Figure 8. Comparison of ΔG_{FEP} and ΔG_{EXP}. The calculated binding free energy was computed using Eq. (4). The ΔG_{EXP} values were computed when the half-maximal inhibitory concentration, IC50, was guessed to be equal to the inhibition constant, K_i.

Calculating the binding affinities of the other SARS-CoV-2 inhibitors. The binding free energy of some available SARS-CoV-2 inhibitors to the Mpro was also evaluated using the assessed approaches. The outcomes were described in **Table 2** and Tables S7-S10 and Figures S6-S7 of the SI file.

Although, the inhibitory of these compounds was extracted from cell-culture experiments[51] meaning that drug targets probably differ from the SARS-CoV-2 Mpro such as RNA polymerase, appropriate correlations between calculated results and experimental data were recorded. Therefore, it may be argued that there are many compounds aiming at inhibiting the Mpro. In particular, in good agreement with the evaluation above, Vina adopts the higher correlation, $R_{Vina} = 0.78 \pm 0.23$ compared with the AD4 package, $R_{AD4} = 0.48 \pm 0.23$. The binding poses of these compounds to SARS-CoV-2 Mpro were thus used as initial structures for SMD/MD refined simulations. The Pearson correlations between FEP, LIE, MM-PBSA, and FPL compared with experimental data are $R_{FEP} = 0.70 \pm 0.16$, $R_{LIE} = 0.67 \pm 0.28$, $R_{MM-PBSA} = 0.00 \pm 0.26$, and $R_{FPL} = -0.71 \pm 0.17$, respectively. The MM-PBSA approach is different from the others since it is very weakly correlated with experiments. Moreover, although, the FEP, LIE, and FPL adopted appropriated results, the linear relationship was decreased. The discrepancies occurred since some compounds probably targeting on RNA polymerase rather than the Mpro.[14]

Table 2. The calculated results in comparison with the experimental values of some compounds to SARS-CoV-2.

Nº	Name	ΔG_{Vina}	ΔG_{AD4}	W	ΔG_{LIE}	$\Delta G_{MM-PBSA}$	ΔG_{FEP}	ΔG_{EXP}[a]
1	Bazedoxifene	-7.5	-8.1	47.4 ± 9.6	-11.12 ± 1.02	-5.13 ± 1.6	-5.25 ± 2.47	-7.48[51]
2	Ciclesonide	-7.4	-8.9	55.9 ± 1.9	-12.89 ± 0.47	-11.58 ± 2.8	-9.87 ± 0.40	-7.34[51]
3	Digitoxin	-8.1	-8.1	72.5 ± 5.2	-13.22 ± 0.93	-0.92 ± 2.95	-16.19 ± 3.88	-9.09[51]
4	Favipiravir	-4.9	-4.9	16.0 ± 1.5	-7.23 ± 0.35	-2.31 ± 1.56	-0.99 ± 0.51	-4.52[51]
5	Gilteritinib	-7.5	-8.5	37.7 ± 2.5	-12.15 ± 0.53	-11.55 ± 3.06	-8.02 ± 0.35	-7.08[51]
6	Lopinavir	-6.3	-5.1	41.8 ± 5.1	-12.39 ± 1.94	-9.30 ± 3.95	-4.72 ± 2.92	-6.59[51]
7	Mefloquine	-6.5	-6.5	45.7 ± 3.0	-10.51 ± 0.16	-9.84 ± 0.06	-3.05 ± 1.36	-7.34[51]
8	Mequitazine	-6.6	-7.7	24.6 ± 2.3	-9.33 ± 0.76	-1.98 ± 4.66	-8.88 ± 0.41	-7.03[51]
9	Niclosamide	-6.6	-6.3	41.9 ± 6.0	-10.95 ± 0.57	-8.38 ± 1.95	-8.77 ± 0.40	-8.97[51]
10	Osajin	-7.0	-7.7	27.6 ± 4.4	-11.45 ± 0.14	-15.26 ± 3.00	-4.15 ± 1.14	-7.41[51]
11	Penfluridol	-6.9	-8.0	59.6 ± 0.5	-10.55 ± 0.38	-0.27 ± 3.41	-10.51 ± 1.51	-7.26[51]
12	Phenazopyridine	-6.0	-6.0	23.9 ± 3.2	-9.96 ± 0.56	-5.90 ± 2.98	-3.80 ± 0.58	-6.23[51]
13	Proscillaridin	-7.6	-7.4	57.8 ± 4.5	-11.85 ± 0.38	-7.32 ± 1.33	-14.56 ± 2.65	-7.79[51]
14	Remdesivir	-6.5	-4.5	37.8 ± 3.9	-12.00 ± 0.31	-28.69 ± 2.94	-8.91 ± 5.65	-6.95[51]

[a]The experimental binding free energies were gained based on IC50 value, approximating that the one equals to the inhibition constant K_i. The unit is kcal mol⁻¹.

CONCLUSIONS

In this context, in order to benchmark which is the appropriate free energy approach for probing the binding free energy of inhibitors to the SARS-CoV-2 Mpro, we have carried out both molecular docking and MD simulations. Vina and AD4 were employed for docking imitations. We have initially demonstrated that Vina package is better than AD4 protocol in both predicting the ligand-binding affinity, $R_{Vina} = 0.60 \pm 0.13$, and binding pose of ligands, successful-docking rate is of ca. 67%, to the SARS-CoV-2 Mpro target. Surprisingly, AD4 formed a poorly correlated results with coefficients of $R_{AD4} = 0.47 \pm 0.21$. It should be noted that the poor accuracy of AD4 was also revealed when the docking results were compared with NEMD simulations, $R = 0.36$.[70]

The MD simulations would be then accomplished. FEP approach was indicated that it provided the most accurate results, $R_{FEP} = 0.85 \pm 0.06$, compared with the respective experiments. Interestingly, the LIE and FPL approach also formed good correlation coefficients, $R_{LIE} = 0.73 \pm 0.09$ and $R_{FPL} = -0.74 \pm 0.11$, while using significantly lower computing resources compared to the FEP, respectively. However, an appropriate relaxed simulation, which is similar to prepare input for FEP/LIE/MM-PBSA calculations, was required to reach equilibrium states before FPL was carried out. Because the successful-docking rate only is ca. 67%, the short NPT simulation may not be sufficient to reach equilibrium states. The MM-PBSA method poorly correlates with the experimental data, $R_{MM-PBSA} = 0.32 \pm 0.29$, as agreed as the recent outcomes.[79]

The atomistic simulations also revealed that the vdW interaction rigidly prevails the cou interaction during the binding of inhibitors to the SARS-CoV-2 Mpro. Moreover, The residues *Thr26*, *His41*, *Ser46*, *Asn142*, *Gly143*, *Cys145*, *His164*, *Glu166*, and *Gln189* play essential factors frequently forming NBC and HB to inhibitors.

ASSOCIATED CONTENT

Supporting Information

Supporting Information Available: The material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Email: ngosontung@tdtu.edu.vn

Author Contributions

STN provided idea and performed molecular docking via the Vina, AD4 and LIE, MM-PBSA, and FEP calculations. NMT carried out FPL simulations. STN and THN prepared the first draft. All authors contributed to writing the manuscript.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

REFERENCES

1.   WHO Coronavirus disease 2019 (COVID-19) Situation Report - 52.

2.   Huang, C. L.; Wang, Y. M.; Li, X. W.; Ren, L. L.; Zhao, J. P.; Hu, Y.; Zhang, L.; Fan, G. H.; Xu, J. Y.; Gu, X. Y.; Cheng, Z. S.; Yu, T.; Xia, J. A.; Wei, Y.; Wu, W. J.; Xie, X. L.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J. G.; Wang, G. F.; Jiang, R. M.; Gao, Z. C.; Jin, Q.; Wang, J. W.; Cao, B., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020,** *395* (10223), 497-506.

3.   Wang, C.; Horby, P. W.; Hayden, F. G.; Gao, G. F., A novel coronavirus outbreak of global health concern. *Lancet* **2020,** *395* (10223), 470-473.

4.   Yu Wai, C.; Chin-Pang, Y.; Kwok-Yin, W., Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like Protease (3CLpro) Structure: Virtual Screening Reveals Velpatasvir, Ledipasvir, and Other Drug Repurposing Candidates. *F1000Res* **2020,** *9*, 129.

5.   Chan, J. F. W.; Yuan, S. F.; Kok, K. H.; To, K. K. W.; Chu, H.; Yang, J.; Xing, F. F.; Liu, J. L.; Yip, C. C. Y.; Poon, R. W. S.; Tsoi, H. W.; Lo, S. K. F.; Chan, K. H.; Poon, V. K. M.; Chan, W. M.; Ip, J. D.; Cai, J. P.; Cheng, V. C. C.; Chen, H. L.; Hui, C. K. M.; Yuen, K. Y., A Familial Cluster of Pneumonia Associated with the 2019 Novel Coronavirus Indicating Person-to-Person Transmission: a Study of a Family Cluster. *Lancet* **2020,** *395* (10223), 514-523.

6.   van Doremalen, N.; Bushmaker, T.; Morris, D. H.; Holbrook, M. G.; Gamble, A.; Williamson, B. N.; Tamin, A.; Harcourt, J. L.; Thornburg, N. J.; Gerber, S. I.; Lloyd-Smith, J. O.; de Wit, E.; Munster, V. J., Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *N Engl J Med* **2020,** *382*, 1564-1567.

7.   FDA Approves First Treatment for COVID-19. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19?fbclid=IwAR3jElh3p4H0YrLtL0o92OE931R6ygixc2edh3zPX4E6SL4AbmMFNu19q8U (accessed Oct 27).

8.   Cohen, J.; Kupferschmidt, K., The 'very, very bad look' of remdesivir, the first FDA-approved COVID-19 drug. *Sci. News* 2020.

9.   Holshue, M. L.; DeBolt, C.; Lindquist, S.; Lofy, K. H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A.; Diaz, G.; Cohn, A.; Fox, L.; Patel, A.; Gerber, S. I.; Kim, L.; Tong, S.; Lu, X.; Lindstrom, S.; Pallansch, M. A.; Weldon, W. C.; Biggs, H. M.; Uyeki, T. M.; Pillai, S. K., First Case of 2019 Novel Coronavirus in the United States. *N. Engl. J. Med.* **2020,** *382* (10), 929-936.

10.   Beigel, J. H.; Tomashek, K. M.; Dodd, L. E.; Mehta, A. K.; Zingman, B. S.; Kalil, A. C.; Hohmann, E.; Chu, H. Y.; Luetkemeyer, A.; Kline, S.; Lopez de Castilla, D.; Finberg, R. W.; Dierberg, K.; Tapson, V.; Hsieh, L.; Patterson, T. F.; Paredes, R.; Sweeney, D. A.; Short, W. R.; Touloumi, G.; Lye, D. C.; Ohmagari, N.; Oh, M.-d.; Ruiz-Palacios, G. M.; Benfield, T.; Fätkenheuer, G.; Kortepeter, M. G.; Atmar, R. L.; Creech, C. B.; Lundgren, J.; Babiker, A. G.; Pett, S.; Neaton, J. D.; Burgess, T. H.; Bonnett, T.; Green, M.; Makowski, M.; Osinusi, A.; Nayak, S.; Lane, H. C., Remdesivir for the Treatment of Covid-19 — Final Report. *N. Engl. J. Med.* **2020**.

11.   Schoeman, D.; Fielding, B. C., Coronavirus envelope protein: current knowledge. *Virology* **2019,** *16* (1), 69.

12.   Fauquet, C. M.; Fargette, D., International Committee on Taxonomy of Viruses and the 3,142 unassigned species. *Virology* **2005,** *2* (1), 64.

13.   Alex, Z.; Vladimir, A.; Alexander, Z.; Bogdan, Z.; Victor, T.; Dmitry S., B.; Daniil, P.; Rim, S.; Andrey, F.; Philipp, O.; Yilin, Y.; Olga, P.; Quentin, V.; Alex, A.; Yan, I., *Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches*. 2020.

14.   Gao, K.; Nguyen, D. D.; Chen, J.; Wang, R.; Wei, G.-W., Repositioning of 8565 Existing Drugs for COVID-19. *J. Phys. Chem. Lett* **2020,** *11* (13), 5373-5382.

15.   Weston, S.; Coleman, C. M.; Haupt, R.; Logue, J.; Matthews, K.; Li, Y.; Reyes, H. M.; Weiss, S. R.; Frieman, M. B., Broad anti-coronaviral activity of FDA approved drugs against SARS-CoV-2 *in vitro* and SARS-CoV *in vivo*. *J. Virol.* **2020**, JVI.01218-20.

16.   Ngo, S. T.; Quynh Anh Pham, N.; Thi Le, L.; Pham, D.-H.; Vu, V. V., Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2020**.

17.   Pham, M. Q.; Vu, K. B.; Han Pham, T. N.; Thuy Huong, L. T.; Tran, L. H.; Tung, N. T.; Vu, V. V.; Nguyen, T. H.; Ngo, S. T., Rapid prediction of possible inhibitors for SARS-CoV-2 main protease using docking and FPL simulations. *RSC Adv* **2020,** *10* (53), 31991-31996.

18.   Ngo, S. T.; Hung Minh, N.; Le Thi Thuy, H.; Pham Minh, Q.; Vi Khanh, T.; Nguyen Thanh, T.; Van, V., Assessing Potential Inhibitors for SARS-CoV-2 Main Protease from Available Drugs using Free Energy Perturbation Simulations. *RSC Adv* **2020,** *10*, 40284-40290.

19.   Marshall, G. R., Computer-Aided Drug Design. *Ann. Rev. Pharmacol. Toxicol.* **1987,** *27*, 193-213.

20.   Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H., Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *J. Chem. Theory Comput.* **2014,** *10* (8), 3331-3344.

21.   Yu, W.; MacKerell, A. D., Computer-Aided Drug Design Methods. In *Antibiotics: Methods and Protocols*, Sass, P., Ed. Springer New York: New York, NY, 2017; pp 85-106.

22.   Ngo, S. T.; Nguyen, T. H.; Tung, N. T.; Nam, P. C.; Vu, K. B.; Vu, V. V., Oversampling Free Energy Perturbation
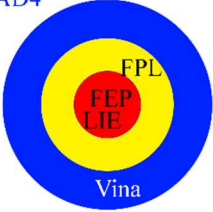
Simulation in Determination of the Ligand-Binding Free Energy. *J. Comput. Chem* **2020,** *41* (7), 611-618.

23.   Decherchi, S.; Cavalli, A., Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.* **2020.**

24.   Yang, J.; Chen, J., *QSAR Analysis of Purine-Type and Propafenone-Type Substrates of P-Glycoprotein Targeting β-Amyloid Clearance* 2013.

25.   Trott, O.; Olson, A. J., Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010,** *31*, 455-461.

26.   Batra, R.; Chan, H.; Kamath, G.; Ramprasad, R.; Cherukara, M. J.; Sankaranarayanan, S. K. R. S., Screening of Therapeutic Agents for COVID-19 Using Machine Learning and Ensemble Docking Studies. *The Journal of Physical Chemistry Letters* **2020,** *11* (17), 7058-7065.

27.   Hansson, T.; Marelius, J.; Åqvist, J., Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aid. Mol. Des.* **1998,** *12* (1), 27-35.

28.   Ngo, S. T.; Hung, H. M.; Nguyen, M. T., Fast and Accurate Determination of the Relative Binding Affinities of Small Compounds to HIV-1 Protease using Non-Equilibrium Work. *J. Comput. Chem.* **2016,** *37* (31), 2734-2742.

29.   Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A., Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *J. Am. Chem. Soc.* **1998,** *120* (37), 9401-9409.

30.   Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; T.E. Cheatham, I., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33* (12), 889-897.

31.   Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys Rev E* **1997,** *56* (5), 5018-5035.

32.   Kollman, P., Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993,** *93* (7), 2395-2417.

33.   Zwanzig, R. W., High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954,** *22* (8), 1420-1426.

34.   Nguyen, N. T.; Nguyen, T. H.; Pham, T. N. H.; Huy, N. T.; Bay, M. V.; Pham, M. Q.; Nam, P. C.; Vu, V. V.; Ngo, S. T., Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *J. Chem. Inf.Model.* **2020,** *60* (1), 204-211.

35.   Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015,** *10* (5), 449-461.

36.   Kuhn, B.; Kollman, P. A., Binding of a siverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000,** *43* (20), 3786-3791.

37.   Schutz, C. N.; Warshel, A., What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **2001,** *44* (4), 400-417.

38.   Hou, T.; Guo, S.; Xu, X., Predictions of binding of a diverse set of ligands to Gelatinase-A by a combination of molecular dynamics and continuum solvent models. *J. Phys. Chem. B* **2002,** *106* (21), 5527-5535.

39.   Genheden, S.; Ryde, U., Comparison of end-point continuum-solvation methods for the calculation of protein-ligand binding free energies. *Proteins* **2012,** *80* (5), 1326-1342.

40.   Uciechowska, U.; Schemies, J.; Scharfe, M.; Lawson, M.; Wichapong, K.; Jung, M.; Sippl, W., Binding free energy calculations and biological testing of novel thiobarbiturates as inhibitors of the human NAD+ dependent histone deacetylase Sirt2. *Med. Chem. Commun.* **2012,** *3* (2), 167-173.

41.   Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.

42.   Andi, B.; Kumaran, D.; Kreitler, D. F.; Soares, A. S.; Shi, W.; Jakoncic, J.; Fuchs, M. R.; Keereetaweep, J.; Shanklin, J.; McSweeney, S. Hepatitis C Virus NSP3/NSP4A Inhibitors as Promising Lead Compounds for the Design of New Covalent Inhibitors for SARS-CoV-2 3CLpro/Mpro Protease. (accessed Oct 04).

43.   Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H., PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016,** *44* (D1), D1202-D1213.

44.   Rathnayake, A. D.; Zheng, J.; Kim, Y.; Perera, K. D.; Mackin, S.; Meyerholz, D. K.; Kashipathy, M. M.; Battaile, K. P.; Lovell, S.; Perlman, S.; Groutas, W. C.; Chang, K.-O., 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV–infected mice. *Sci. Transl. Med.* **2020,** *12* (557), eabc5332.

45.   Dai, W.; Zhang, B.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L.-K.; Xu, Y.; Yang, H.; Liu, H., Structure-based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020,** *368*, 1331-1335.

46.   Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R., Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* **2020,** *368*, 409-412.

47.   Su, H.-x.; Yao, S.; Zhao, W.-f.; Li, M.-j.; Liu, J.; Shang, W.-j.; Xie, H.; Ke, C.-q.; Hu, H.-c.; Gao, M.-n.; Yu, K.-q.; Liu, H.; Shen, J.-s.; Tang, W.; Zhang, L.-k.; Xiao, G.-f.; Ni, L.; Wang, D.-w.; Zuo, J.-p.; Jiang, H.-l.; Bai, F.; Wu, Y.; Ye, Y.; Xu, Y.-c., Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol. Sin.* **2020,** *41* (9), 1167-1177.

48.   Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J., Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res* **2020,** *30* (8), 678-692.

49.   Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H., Structure of Mpro from SARS-CoV-2 and Discovery of its Inhibitors. *Nature* **2020,** *582* (7811), 289-293.

50.   Vuong, W.; Khan, M. B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H. A.; McKay, R. T.; van Belkum, M. J.; Joyce, M. A.; Young, H. S.; Tyrrell, D. L.; Vederas, J. C.; Lemieux, M. J., Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nature Communications* **2020,** *11* (1), 4282.

51.   Jeon, S.; Ko, M.; Lee, J.; Choi, I.; Byun, S. Y.; Park, S.; Shum, D.; Kim, S., Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs. *Antimicrob. Agents Chemother.* **2020,** *64* (7), e00819-20.

52.   Ngo, S. T.; Quynh Anh Pham, N.; Thi Le, L.; Pham, D.-H.; Vu, V. V., Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease. *Journal of Chemical Information and Modeling* **2020**, *In press*.

53.   Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19-25.

54.   Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M., Motional Timescale Predictions by Molecular Dynamics Simulations: Case Study using Proline and Hydroxyproline Sidechain Dynamics. *Proteins: Struct., Funct., Bioinf.* **2014**, *82* (2), 195-215.

55.   Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157-1174.

56.   Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E. C., III, V.W.D. ; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; SalomonFerrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; D.M., Y.; P.A., a. K., AMBER 18. *University of California, San Francisco* **2018**.

57.   Sousa da Silva, A. W.; Vranken, W. F., ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **2012**, *5* (1), 1-8.

58.   Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089-10092.

59.   Park, S.; Schulten, K., Calculating Potentials of Mean Force from Steered Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *120* (13), 5946-5961.

60.   Almlöf, M.; Brandsdal, B. O.; Åqvist, J., Binding affinity prediction with different force fields: Examination of the linear interaction energy method. *J. Comput. Chem.* **2004**, *25* (10), 1242-1254.

61.   Shrake, A.; Rupley, J. A., Environment and exposure to solvent of protein atoms-lysozyme and insulin. *J. Mol. Biol.* **1973**, *79* (2), 351-371.

62.   Sitkoff, D.; Sharp, K. A.; Honig, B., Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *97* (7), 1978-1988.

63.   Sharp, K. A.; Honig, B., Electrostatic interactions in macromolecules: theory and applications. Annu. Rev. Biophys. . *Biophys. Chem.* **1990**, *19*, 301-332.

64.   Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (18), 10037-10041.

65.   McQuarrie, D. A., *Statistical Thermodynamics*. Harper and Row: New York, 1973.

66.   Ytreberg, F. M., Absolute FKBP Binding Affinities Obtained via Non-Equilibrium Unbinding Simulations. *J. Chem. Phys.* **2009**, *130* (16), 164906.

67.   Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245-268.

68.   Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research* **2007**, *35* (suppl_2), W522-W525.

69.   Efron, B., Bootstrap Methods: Another Kook at the Jackknife. *Ann. Stat.* **1979**, *7*, 1-26.

70.   Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P., Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *J. Chem. Theor. Comput.* **2020**, *16* (11), 7160-7172.

71.   Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.; Lorenzen, K.; Krichel, B.; Kopicki, J.-D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Pérez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hänle, A.; Groessler, M.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schlünzen, F.; Xavier, P. L.; Werner, N.; Andaleeb, H.; Ullah, N.; Falke, S.; Srinivasan, V.; Franca, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitsev-Doyle, J. J.; Knoska, J.; Peña Murillo, G. E.; Mashhour, A. R.; Guicking, F.; Hennicke, V.; Fischer, P.; Hakanpää, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; Stetten, D. v.; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia Alai, M. M.; Niebling, S.; Günther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E.-C.; Mehrabi, P.; Karničar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Uetrecht, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Günther, S.; Turk, D.; Hinrichs, W.; Chapman, H. N.; Pearson, A. R.; Betzel, C.; Meents, A., Inhibition of SARS-CoV-2 Main Protease by Allosteric Drug-Binding. *bioRxiv* **2020**, 2020.11.12.378422.

72.   Bjelic, S.; Nervall, M.; Gutiérrez-de-Terán, H.; Ersmark, K.; Hallberg, A.; Åqvist, J., Computational inhibitor design against malaria plasmepsins. *Cell. Mol. Life Sci.* **2007**, *64* (17), 2285-2305.

73.   Ngo, S. T.; Mai, B. K.; Derreumaux, P.; Vu, V. V., Adequate prediction for inhibitor affinity of Aβ40 protofibril using the linear interaction energy method. *RSC Adv* **2019**, *9* (22), 12455-12461.

74.   Zhang, H.; Yin, C.; Jiang, Y.; van der Spoel, D., Force Field Benchmark of Amino Acids: I. Hydration and Diffusion in Different Water Models. *J. Chem. Inf. Model.* **2018**, *58* (5), 1037-1052.

75.   Zhang, H.; Jiang, Y.; Cui, Z.; Yin, C., Force Field Benchmark of Amino Acids. 2. Partition Coefficients between Water and Organic Solvents. *J. Chem. Inf. Model.* **2018**, *58* (8), 1669-1681.

76.   Viet, M. H.; Ngo, S. T.; Lam, N. S.; Li, M. S., Inhibition of Aggregation of Amyloid Peptides by Beta-Sheet Breaker Peptides and Their Binding Affinity. *J. Phys. Chem. B* **2011**, *115* (22), 7433-7446.

77.   Ngo, S. T.; Li, M. S., Curcumin Binds to Aβ$_{1-40}$ Peptides and Fibrils Stronger than Ibuprofen and Naproxen. *J. Phys. Chem. B* **2012**, *116* (34), 10165-75.

78.   Ngo, S. T.; Hung, H. M.; Truong, D. T.; Nguyen, M. T., Replica Exchange Molecular Dynamics Study of the Truncated Amyloid Beta (11-40) Trimer in Solution. *Phys. Chem. Chem. Phys.* **2017**, *19* (3), 1909-1919.

79.   Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.; Zhou, L.; Lin, Y.; Wu, D.; Zhang, L.; Liu, H.; Xu, X.; Yu, K.; Zhang, Y.; Cui, J.; Zhan, C.-G.; Wang, X.; Luo, H.-B., Identify Potent SARS-CoV-2 Main Protease Inhibitors via Accelerated Free Energy Perturbation-Based Virtual Screening of Existing Drugs. *Proc. Natl. Acad. Sci. U.S.A* **2020**, *117* (44), 27381-27387.

80.   Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011,** *51* (1), 69-82.

81.   Ryde, U.; Soderhjelm, P., Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* **2016,** *116* (9), 5520-5566.

AD4

FPL

FEP
LIE

Vina

MM-PBSA